



2010 Committee on Korean Materials (CKM) Annual Meeting

**Robert Miller
Director of Books
Internet Archive
March 25, 2010**



The Asami Collection:

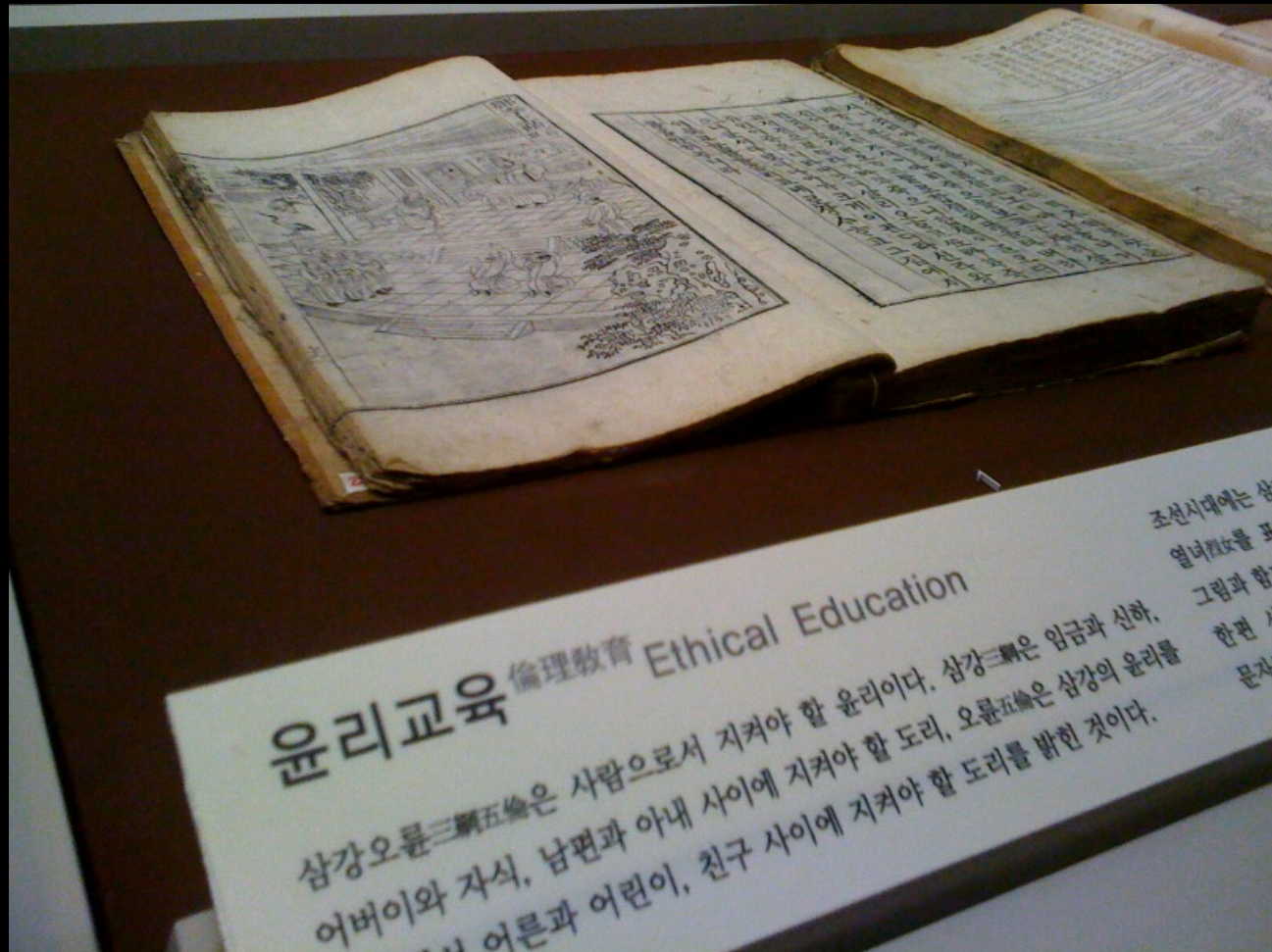
**“Finding Innovation
between the pages!**

...Or...

**Preserving, Transforming
and Accessing”**



Korean printing experience





Korean printing experience





Korean printing experience

한글의 발명가인 세종대왕의 생애와 한글의 발달 과정

King *Sejong* (1418~1450), the fourth king of *Joseon*, who felt compassion toward the common people, invented the unique Korean alphabet, *Hangeul*, using phonetic symbols in December of the twenty-fifth year of his reign and promulgated *Hunminjeongeum* in 1446. *Hangeul* consists of seventeen consonants and eleven vowels. It is the only language in the world with a known inventor and time of creation. Even though many people severely opposed the initial promulgation of *Hangeul*, King *Sejong* aggressively encouraged its use by writing the first book in *Hangeul*, *Yongbieocheonga*. *Hangeul* made it possible for people to be able to write their thoughts and ideas, which led to the improvement of living conditions and the growth of consciousness. It was used for resolving civil petitions, disseminating agricultural techniques, and exchanging letters with friends and relatives.

my job today



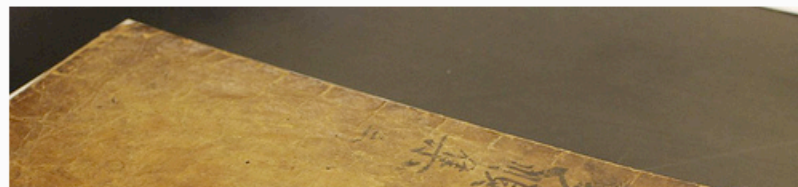
Today we looked at 600+ year-old book from Korea that has a stain. We're using SCIENCE to help the conservator figure out what the stain is, and if it's damaging the paper more than it already has, so she can treat it and house it safely. It's beautiful and soft and makes the best tissuey sounds when we turn the pages.

Comments



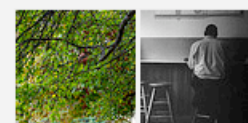
[volcanojw](#) pro says:

the cover:



Uploaded on November 4, 2009
by [volcanojw](#)

volcanojw's photostream

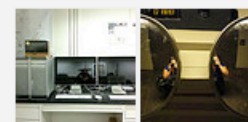


2,548 uploads

browse

This photo also belongs to:

LC Wanderings (Set)



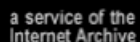
96 items

browse

2 people call this photo a favorite

Tags

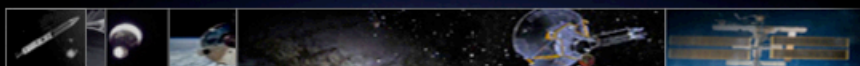
- Korean
- book
- 15th century
- moveable type
- paper
- stained
- trying to help it
- using science
- conservation
- LC



Shepard Completes His Mission



Astronauts



Spaceflight Timeline

Follow us on [Twitter](#)

Visit [The NASA Images Shop](#) and [The Space Store](#) to support us.

[About](#) | [Terms](#) | [Contact](#) | [Powered by Luna Imaging](#)

The Internet Archive is a 501(c)(3) non-profit organization.

Search by Keyword

Featured Collections



**Videographer
of the Year**

[View All Collections...](#)

Visit NASA.gov



The access...

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome | Login | Register

Search LUNA

Search

Advanced Search

◀ ▶ Collections Explore Create Share This Help

Browse the Asami Collection and Korean Rare Books

Collection Selection

• All Collections (44)

LUNATest Collections

• Asami Collection and Korean Rare Books

Featured Media



Category Pages

Browse All

The Asami collection of classical Korean literature is comprised of printed books, manuscripts, and rubbings, including numerous examples of early movable-type printing, some dating as early as the seventeenth century. Aside from its artifactual significance, the collection is a valuable resource for the study of the political and legal institutions of Yi dynasty Korea. The C. V. Starr East Asian Library at the University of California, Berkeley possesses other Korean rare materials. In all, the Starr's rare Korean holdings total some 2,400 titles in approximately 1.2 million pages. The Asami Collection and Korean Rare Books project was launched by the C. V. Starr East Asian Library and the Institute of Korean Culture, Korea University. Texts and images were scanned by Internet Archive and put online by David Rumsey. Please direct all comments, suggestions, and queries to the C. V. Starr



Asami Collection and
Korean Rare Books

C. V. Starr East Asian Library,
University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Search

Advanced Search

◀ ▶ Collections Explore Create Share This Help

Login

Welcome to LUNA

Please login

Username:

Password:

Submit

[Have you forgotten your password?](#)

<http://luna.davidrumsey.com:9080/luna/servlet/>

A Google search bar with the word "Google" and a magnifying glass icon.A Google search bar with the word "Google" and a magnifying glass icon.

100%

• Asami Collection and Korean Rare Books

[Browse All](#)

Copyright Information
The Asami collection and all Korean rare holdings at the Starr Library
are copyrighted by the Regents of the University of California on behalf
of the C. V. Starr East Asian Library, 2009, all rights reserved. Images
may not be reproduced without permission.
Contact the Starr Library at eal@library.berkeley.edu.

The Asami collection and all Korean rare holdings at the Starr Library are copyrighted by the Regents of the University of California on behalf of the C. V. Starr East Asian Library, 2009, all rights reserved. Images may not be reproduced without permission.
Contact the Starr Library at eal@library.berkeley.edu.

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Advanced Search

Search

◀ ▶ Collections **Explore** Create Share This Embed This Help

Narrow Search

What

Show More

- Manuscripts, K ... (44)
- Kyongbu (30)
- Yi jing (16)
- Sabu (14)
- Shu jing (6)

Who

- C. V. Starr Ea ... (44)
- 880-01 Chong, ... (44)

When

Show More

- 18--?] (27)
- 1808 (12)
- Sunjo 8 (12)
- 18--? (3)
- 1817 (2)

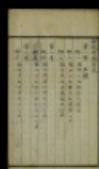
Advanced Search

Browse All

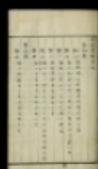
1-44 of 44



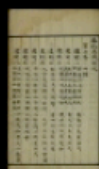
Printer Friendly



Chunchu kojing ...
1
880-01 Chong, Y ...
18--?]



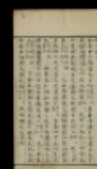
Chunchu kojing ...
2
880-01 Chong, Y ...
18--?]



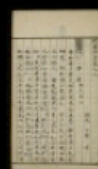
Chunchu kojing ...
3
880-01 Chong, Y ...
18--?]



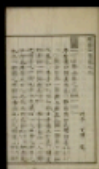
Chunchu kojing ...
4
880-01 Chong, Y ...
18--?]



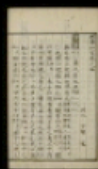
Chuyok sajon : ...
1
880-01 Chong, Y ...
1808, Sunjo 8



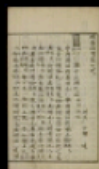
Chuyok sajon : ...
2
880-01 Chong, Y ...
1808, Sunjo 8



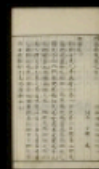
Chuyok sajon : ...
3
880-01 Chong, Y ...
1808, Sunjo 8



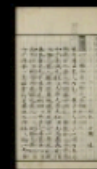
Chuyok sajon : ...
4
880-01 Chong, Y ...
1808, Sunjo 8



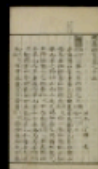
Chuyok sajon : ...
5
880-01 Chong, Y ...
1808, Sunjo 8



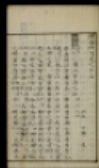
Chuyok sajon : ...
6
880-01 Chong, Y ...
1808, Sunjo 8



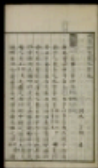
Chuyok sajon : ...
7
880-01 Chong, Y ...
1808, Sunjo 8



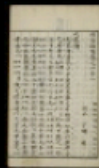
Chuyok sajon : ...
8
880-01 Chong, Y ...
1808, Sunjo 8



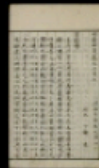
Chuyok sajon : ...
9



Chuyok sajon : ...
10



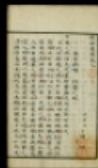
Chuyok sajon : ...
11



Chuyok sajon : ...
12



Kyongse yu
1



Kyongse yupyo
2

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Advanced Search

Search

◀ ▶ Collections Explore Create Share This Embed This Help

Narrow Search

What

Show More

- Manuscripts, K ... (44)
- Kyongbu (30)
- Yi jing (16)
- Sabu (14)
- Shu jing (6)

Who

- C. V. Starr Ea ... (44)
- 880-01 Chong, ... (44)

When

Show More

- 18--?] (27)
- 1808 (12)
- Sunjo 8 (12)
- 18--? (3)
- 1817 (2)

Advanced Search

Browse All

1-44 of 44



Printer Friendly

Chunchu kojing ... 1
880-01 Chong, Y ... 18--?]

Chunchu kojing ... 4
880-01 Chong, Y ... 18--?]

Chunyok sajong : ... 1
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 2
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 3
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 4
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 5
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 6
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 7
880-01 Chong, Y ... 1808, Sunjo 8

Chunyok sajong : ... 8
880-01 Chong, Y ... 1808, Sunjo 8

Kyongse yu ... 1

Kyongse yupyo ... 2

Chunyok sajong : ... 9

Chunyok sajong : ... 10

Chunyok sajong : ... 11

Chunyok sajong : ... 12

Media Summary

Title: Chunchu kojing : kwon 1-12

Volume: 1

Creator: 880-01 Chong, Yag-yong, 1762-1836

Description: 12-kwon 4-chaek, saju ssangbyon, kwanggwak 18.3 x 12.5 cm., yugy...

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Search

Advanced Search

Collections Explore Create Share This Embed This Help

Media Information

Collection:

Asami Collection and Korean Rare Books

Identifier:

chunchukojingkw001ri ch

Title:

Chunchu kojing : kwon 1-12

Volume:

1

Creator:

880-01 Chong, Yag-yong, 1762-1836

Subject:

Manuscripts, Korean

Subject:

Kyongbu

Subject:

Chun chiu

Description:

Manuscript, on double leaves, oriental style

Description:

12-kwon 4-chaek; saju ssangbyon; kwanggwak 18.3 x 12.5 cm.; yugye; 10-haeng; mu omi

Publisher:

[Korea : s.n.]

Date:

18--?]

Language:

kor

Page Progression:

Chunchu kojing : kwon 1-12

Default Group

Add To Media Group



Zoom: Auto



Chunchu kojing : kwon 1-12



Page: n0



search...

go

Search results



Internet Archive

Content Problems | Report Bugs | Texts Collection | Contact Us

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Search

Advanced Search

Collections Explore Create Share This Embed This Help

Media Information

Collection:

Asami Collection and Korean Rare Books

Identifier:

chunchukojingkw01ri ch

Title:

Chunchu kojing : kwon 1-12

Volume:

1

Creator:

880-01 Chong, Yag-yong, 1762-1836

Subject:

Manuscripts, Korean

Subject:

Kyongbu

Subject:

Chun chiu

Description:

Manuscript, on double leaves, oriental style

Description:

12-kwon 4-chaek; saju ssangbyon; kwanggwak 18.3 x 12.5 cm.; yugye; 10-haeng; mu omi

Publisher:

[Korea : s.n.

Date:

18--?]

Language:

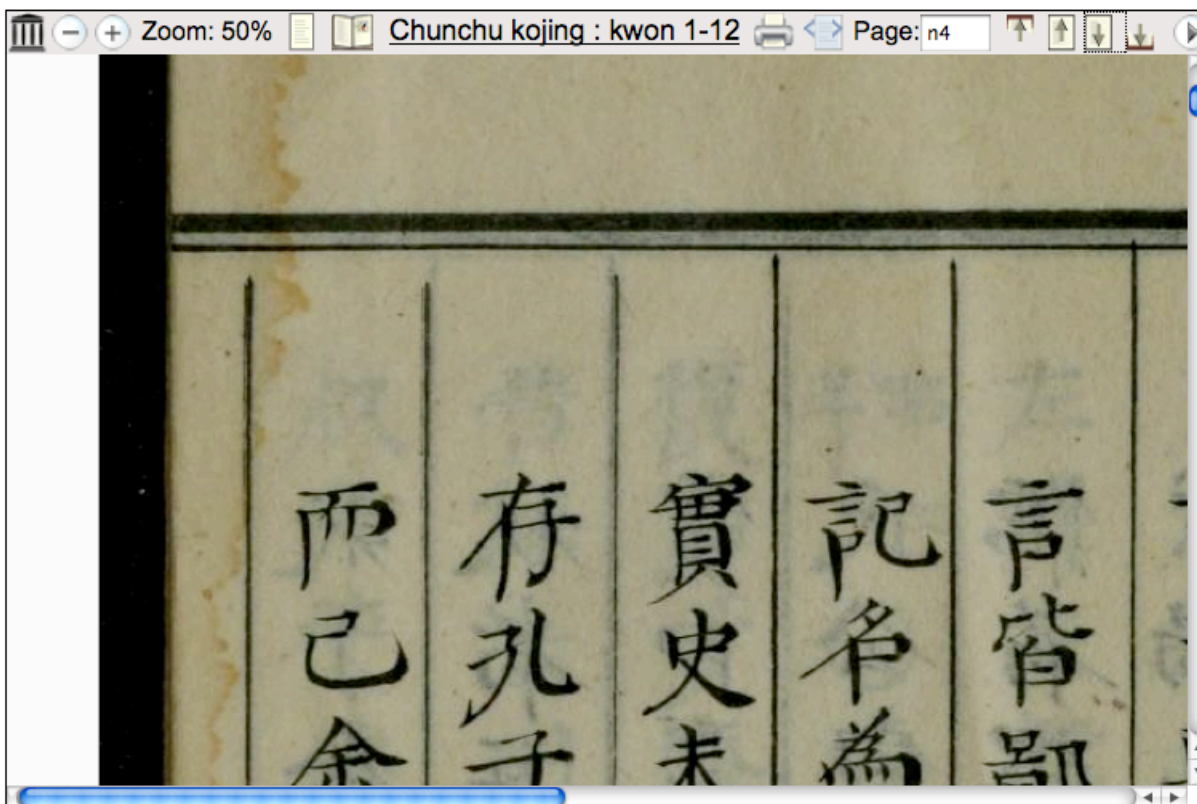
kor

Page Progression:

Chunchu kojing : kwon 1-12

Default Group

Add To Media Group



Internet Archive

Content Problems | Report Bugs | Texts Collection | Contact Us

Reconstru...ON IDEAS

BPL-IA Contract DRAFT 2010011

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: Default Group

Search LUNA

Search

Advanced Search

◀ ▶ Collections Explore **Create** Share This Help

Create New Media Group

Name (max 50 characters):

philly

Publicly Viewable:

☒ Yes ☐ No

Password (optional) 2-8 characters:

korea

Folder:

My Groups

Active Media Group:

☒

Description (optional - max 50 characters):

This is an example of what one may do with a new media group!

Cancel

Save



Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: philly

Search LUNA

Search

Advanced Search

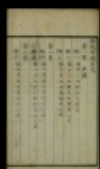
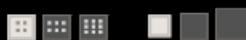
Collections Explore Create Share This Embed This Help

Media Group: philly

Printer Friendly

Properties Manage Group Reorder Copy Slide Show Add External Media

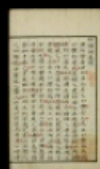
1-3 of 3



Chunchu kojing ...
1
880-01 Chong, Y ...
18--?]



Chunchu kojing ...
4
880-01 Chong, Y ...
18--?]



Kyongse yu
1
880-01 Chong, Y ...
18--?]

1-3 of 3



Asami Collection and
Korean Rare Books

C. V. Starr East Asian Library,
University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: philly

Search LUNA

Advanced Search

Collections Explore Create Share This Help

External Media Search

Find [Flickr](#) images and add them to your [Active Media Group](#) using the search below.

Keyword

Username

asami

AND/OR

Search

1 2 3 ... 16 17

1-30 of 500



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



Go To Source



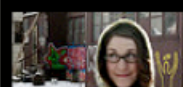
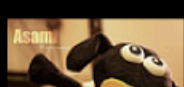
Go To Source



Go To Source



Go To Source



my job today



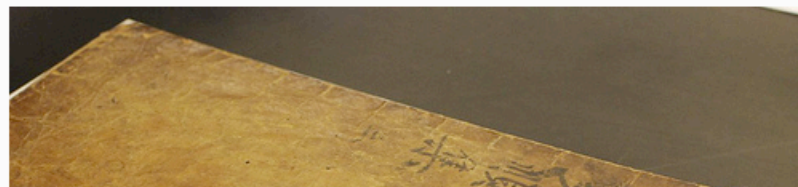
Today we looked at 600+ year-old book from Korea that has a stain. We're using SCIENCE to help the conservator figure out what the stain is, and if it's damaging the paper more than it already has, so she can treat it and house it safely. It's beautiful and soft and makes the best tissuey sounds when we turn the pages.

Comments



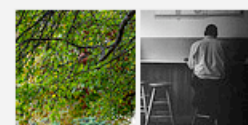
[volcanojw](#) pro says:

the cover:



Uploaded on November 4, 2009
by [volcanojw](#)

volcanojw's photostream

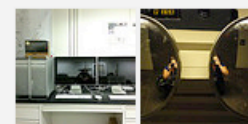


2,548 uploads

browse

This photo also belongs to:

LC Wanderings (Set)



96 items

browse

2 people call this photo a favorite

Tags

- Korean
- book
- 15th century
- moveable type
- paper
- stained
- trying to help it
- using science
- conservation
- LC

Asami Collection and Korean Rare Books | C. V. Starr East Asian Library, University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: philly

Search LUNA

Search

Advanced Search

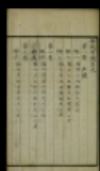
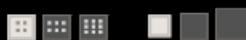
Collections Explore Create Share This Embed This Help

Media Group: philly

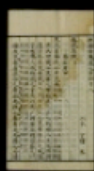
Printer Friendly

Properties Manage Group Reorder Copy Slide Show Add External Media

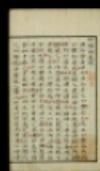
1-4 of 4



Chunchu kojing ...
1
880-01 Chong, Y ...
18--?]



Chunchu kojing ...
4
880-01 Chong, Y ...
18--?]



Kyongse yu
1
880-01 Chong, Y ...
18--?]



1-4 of 4



Asami Collection and
Korean Rare Books | C. V. Starr East Asian Library,
University of California, Berkeley

Welcome robert-miller | Edit Settings/Defaults | Logout

Active Media Group: philly

Search LUNA

Search

Advanced Search

Collections **Explore** Create Share This Help

My Presentations

[Create Presentation](#) | [Create Folder](#)

▼ Search My Presentations

Keyword

☐ Include Locked

Search

Reset

philly 2

Public Presentations

1 of 2

▼ Search Public Media Groups

Keyword

Username

AND/OR

Search

Reset

☒ Include Locked

philly 2 (robert-miller)

Show More



OCR

- The Korean book, like others in this collection of 1600s-to-mid 1800s Korean books, is in Classical Chinese, then the 'lingua franca' for scholars in the region.
- We can expect many of these types of books to be not movable type, but pages carved as a whole. This particular instance seems to be handwritten.
- The Abby 9 OCR results are then predictably abysmal. Correctly recognized 6 of about 170 total (handwritten) characters on the page.

1 Document

Scan Open

Document Languages
Chinese (PRC); Chinese (Taiwan); Engl...

1 2

2 Image

Read

Analyze Select Recognition Area

Text Picture Table

Delete Separator Vertical Separator Horizontal Separator

Order Delete Edit Image

3 Text

PDF Document (Save Exact copy

B I U x² x₂

學，對
省覺 付 魔 非 覺
要耳 覺 恪馬倅
馬良馬浙 e 謂 夸膝 省目七
攝小 ； 難、入肩亦 一
沈學 尺三間 一 十孕省 索？
&聚 非 K 非及 L 募麗

小引

學者何學也者覺也覺者何覺也者覺其非也覺其非
奈何于雅言覺之爾言之而喚鼠為珠娥而覺之曰是
鼠耳吾安耳言之而指鹿為馬娥而覺之曰是鹿耳吾
安耳既覺而愧焉悔焉改焉斯之謂學學修己者曰勿
以惡小而為之學治文者亦勿以惡小而為之斯其學
有進已處遐遠者學文皆傳聞耳多訛舛故有足言也
然舉一而反三聞一而知十學者之責索言之不能窮
故聚言之非其非止是也嘉慶己卯冬鐵馬山樵書

Hi All,

Just a followup (this time with screenshots) on the OCR test on the Korean books collection from Cal's East Asian Library.

Test book:

http://www.archive.org/details/korean_demo_00_test

We tried Abby 9 OCR on this page:

http://www.us.archive.org/GnuBook/?id=korean_demo_00_test#11

A screenshot of the results is attached. (Robert is interested in these for presentations so folks understand the contours of the current technology.)

The Korean book, like others in this collection of 1800s-to-mid 1900s Korean books, is in Classical Chinese, then the 'lingua franca' for scholars in the region.

We can expect many of these types of books to be not movable type, but pages carved as a whole. This particular instance seems to be handwritten.

The Abby 9 OCR results are then predictably abysmal. Correctly recognized 6 of about 170 total (handwritten) characters on the page.

The two attached screenshots show (1) the results of the page recognition, and (2) the results of the page recognition with correctly recognized characters circled.

The curator of the Korean books special collection notes there is a Korean OCR package available for purchase (if we ever start doing Korean books that are in Korean).

Anna

--

Anna Naruta, PhD
Collections Manager
Internet Archive - archive.org

Archive.org's projects include
OpenLibrary.org

member of
OpenContentAlliance.org

in The Economist 2009 March:
http://www.economist.com/science/tq/displaystory.cfm?story_id=13174399&fsrc=rss
launch with Sun:
<http://www.sun.com/featured-articles/2009-0325/feature/index.jsp>

Presidio of San Francisco
116 Sheridan Ave
San Francisco, CA 94129
naruta@archive.org
(415) 561-6767



The digital preservation ...and.. the transformation



Equipment- Alternate A

Needs-

- Color
- Reduce damage to book
- Output formats





Equipment Alternate B-robotics

Needs-

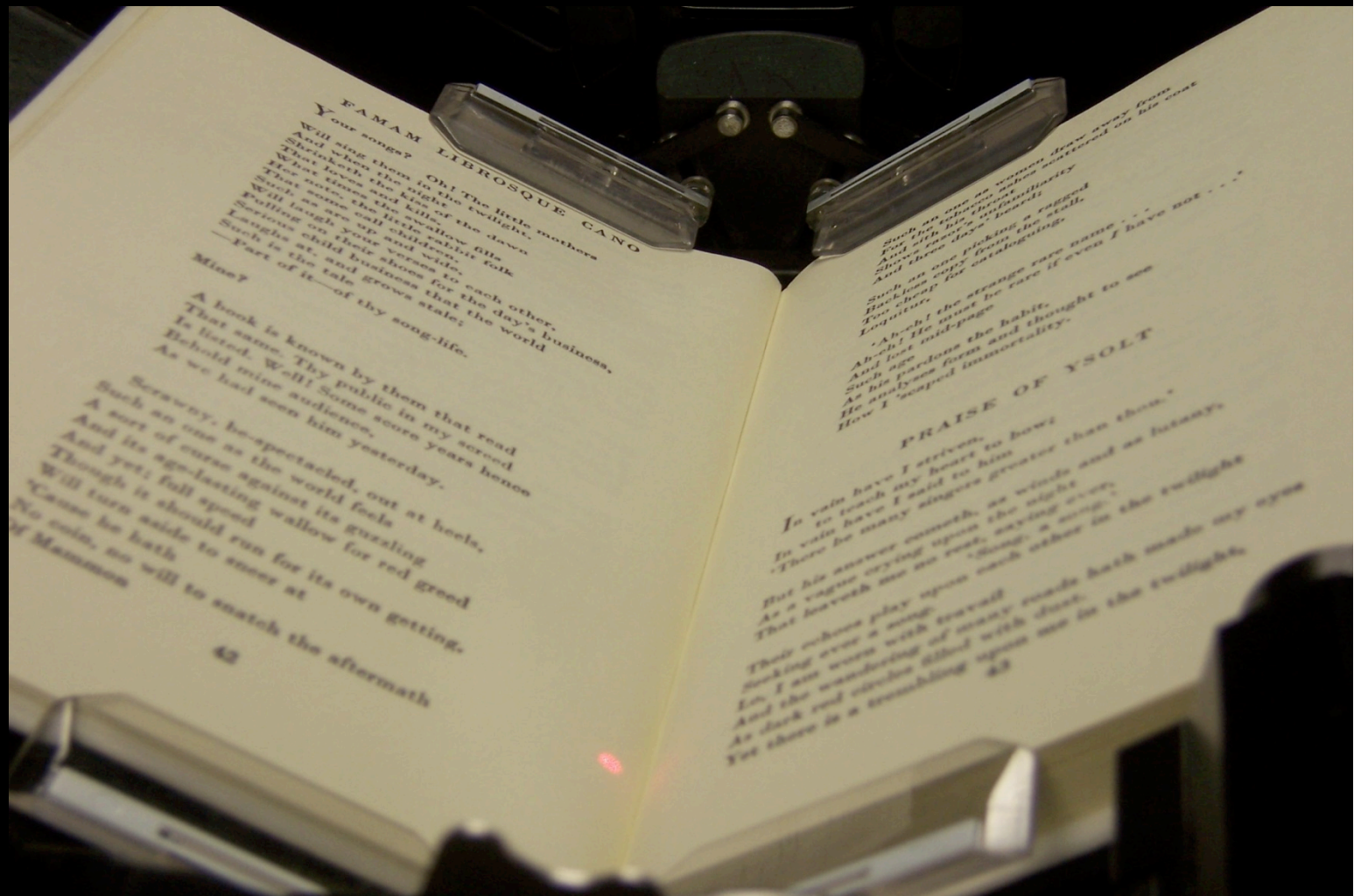
- Low capital cost
- Open source code
- Not just pages per scanner hr
— pages/labor hour



Findings from studies

Observations:

- set up time
- missing pages
- curvature-ocr
- post-processing





Solution- Scribe

Pluses-

- Low cost
- Flexible
- Chassis approach





Floating Cradle and platen



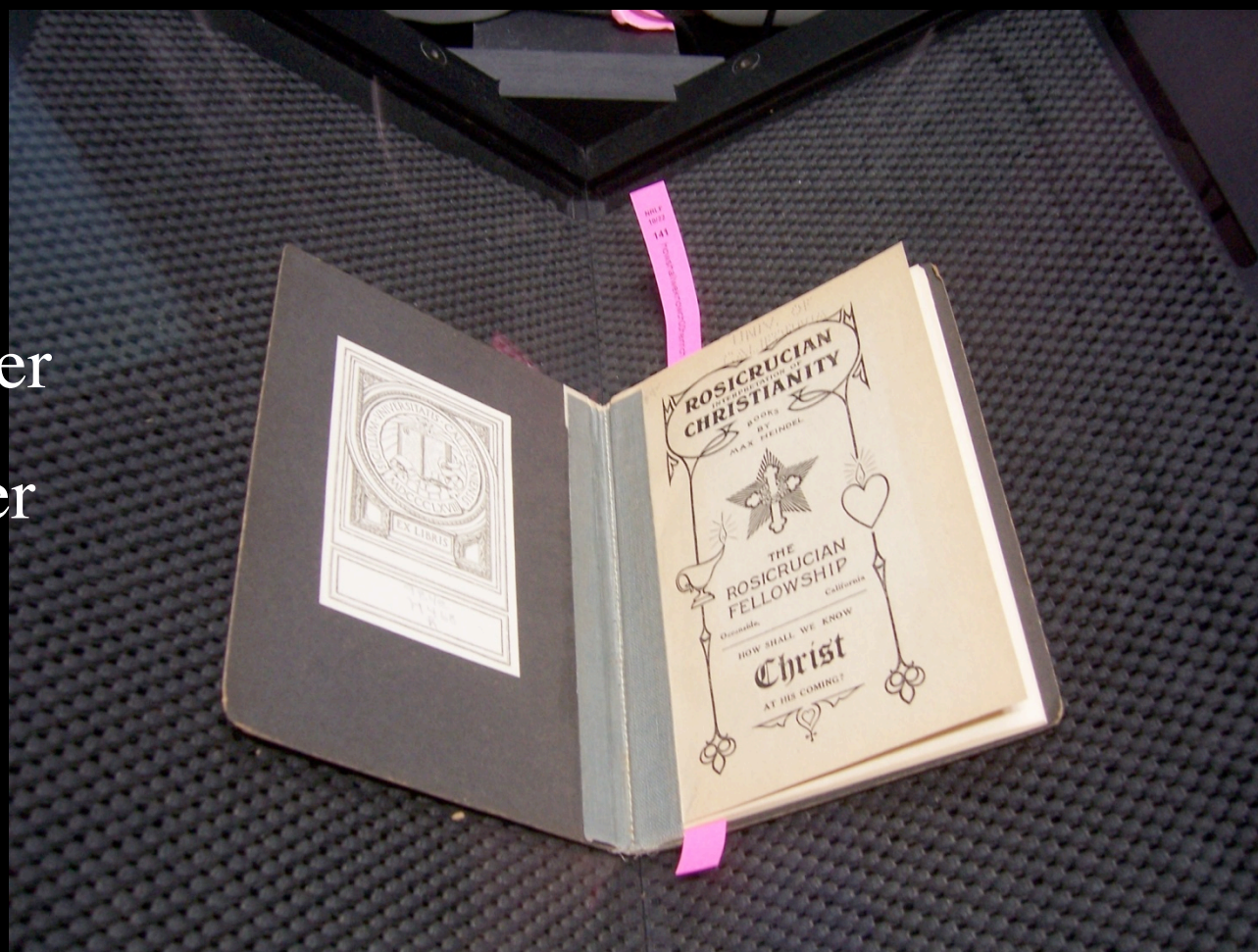


Path of a book not born digital

Goal-

\$0.10 per pg

- Storage forever
- Access forever
- Lifetime mgt of files





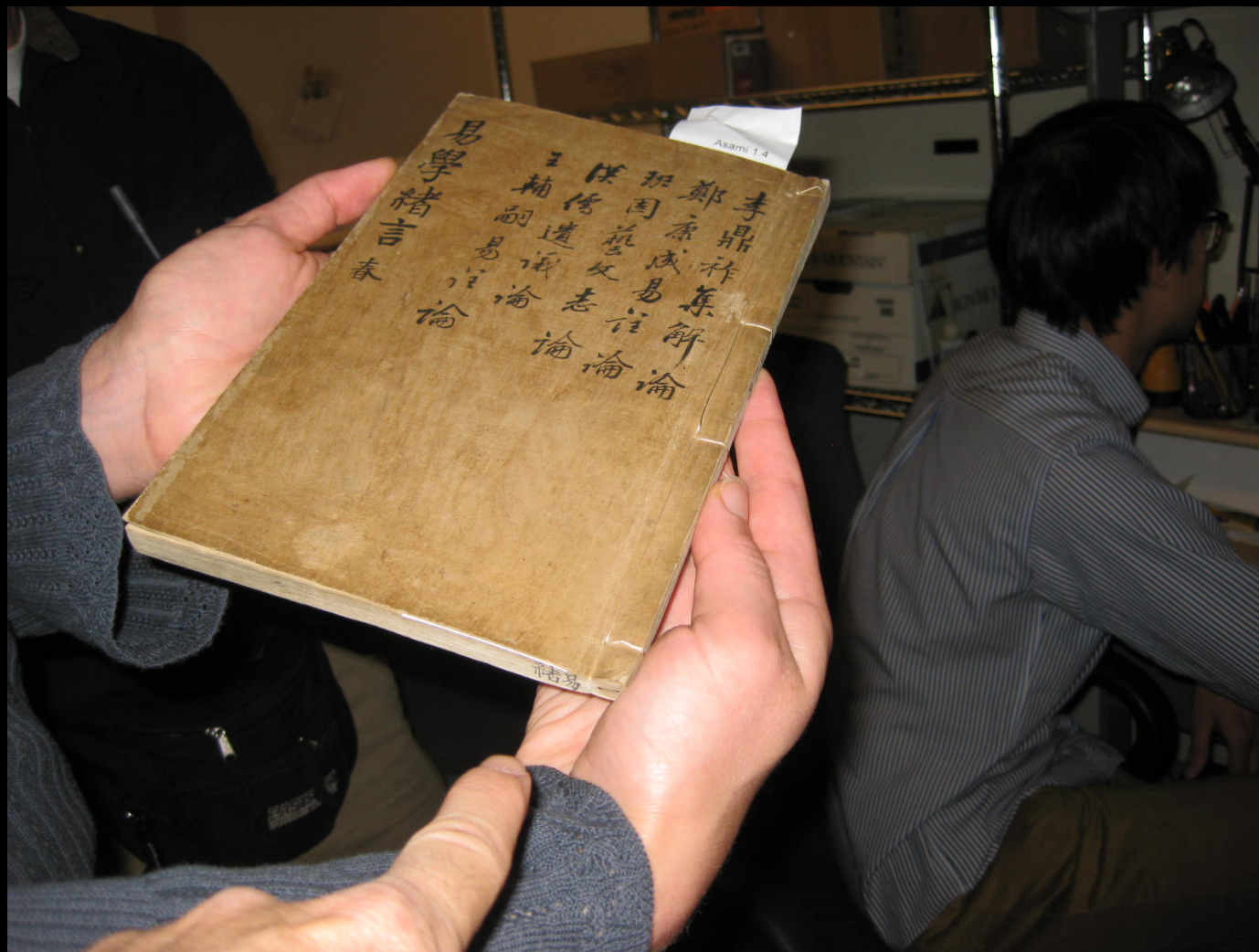
Automated Record Retrieval

- Z39.50
- MARC
- DC
- Persistent identifier
- Bindings
- Gutter
- MULs





Asami- ready to scan





We don't scan flat



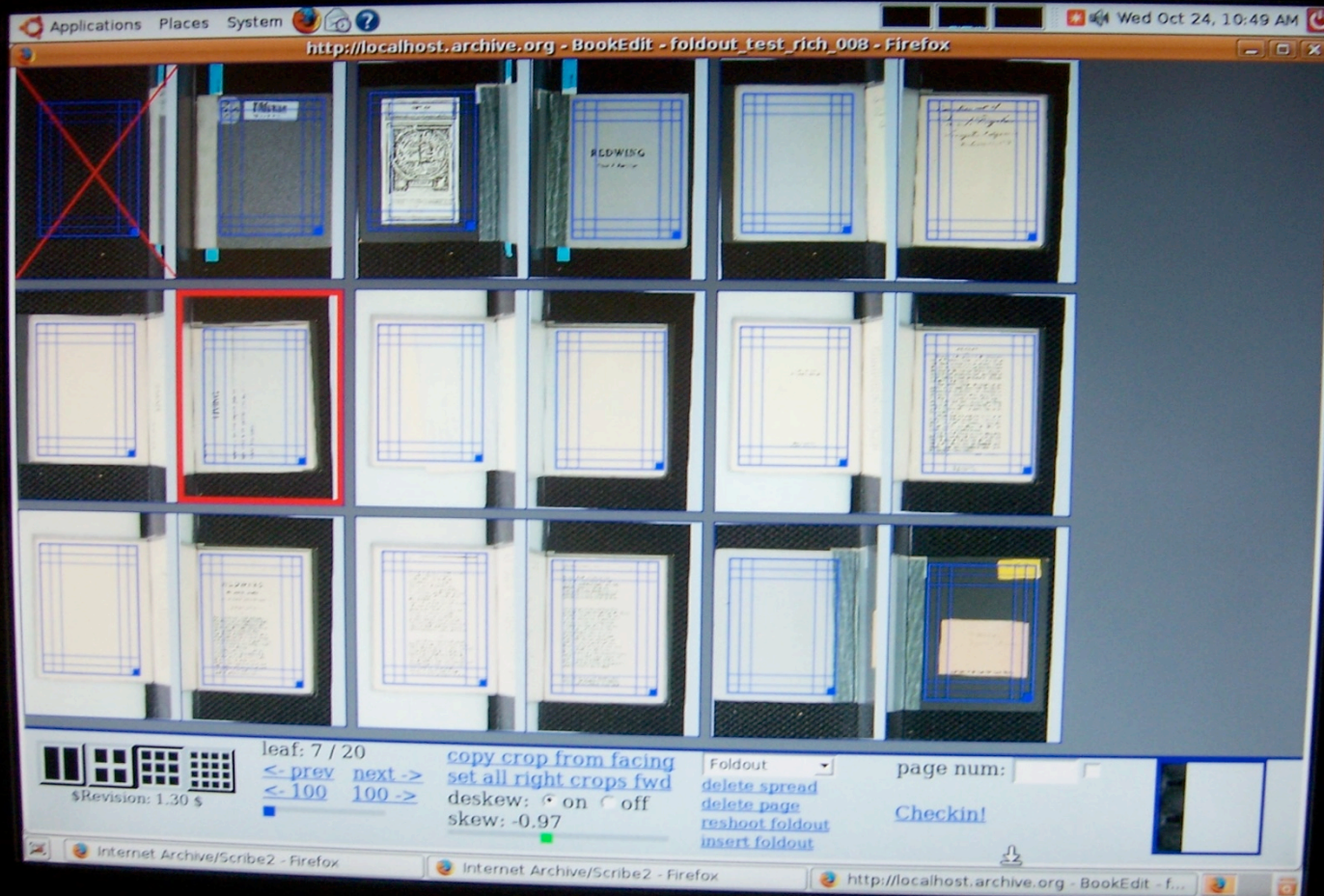


Asami- right-to-left



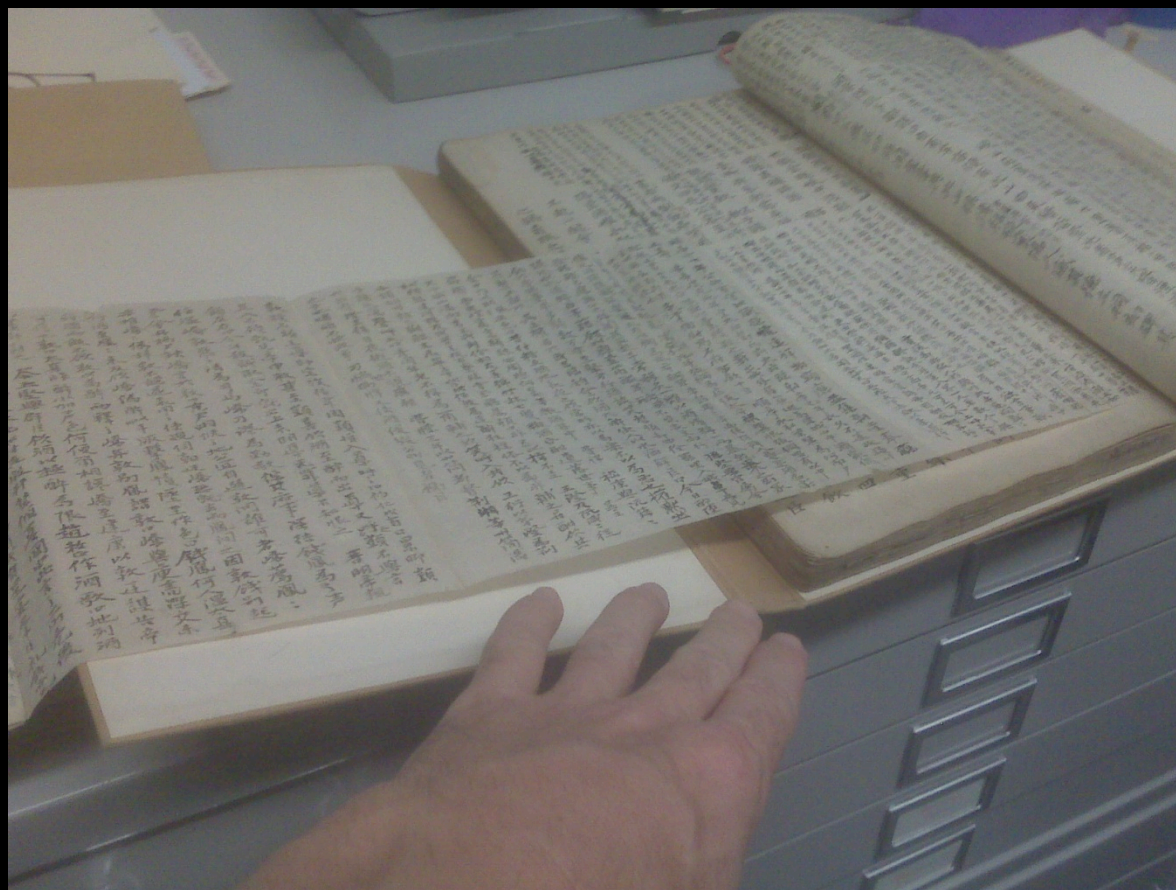
Software tools-

- Deskew
- Structural meta data
- Missing page detection
- Same station post processing





Asami foldouts





Processing, formats and storage

- Storage-tifs→jpg2
- Access- DjVU,
PDF
- OCR- Abby 8.0
- Persistent
identifier
- Lifetime file mgt





Cost

- \$0.10 per page
 - Access and storage files
 - Life time management of files
 - Storage...forever
 - Serve up files or download them
 - [Internet Archive American Libraries goody.htm](#)



10 minute book press





Thank You!

Robert Miller

Director of Books

01 415 640 1092

robert@archive.org